

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
10 September 2004 (10.09.2004)

PCT

(10) International Publication Number
WO 2004/077884 A1

(51) International Patent Classification⁷: **H04S 3/00**,
H04R 1/20 // 3/00

Helsinki (FI). PULKKI, Ville [FI/FI]; Yläportti 4 A 7,
FI-02210 Espoo (FI).

(21) International Application Number:
PCT/FI2004/000093

(74) Agent: KUOSMANEN, Panu; Helsinki University of
Technology, Otaniemi International Innovation Centre
OHC, P.O. Box 9202, FI-02015 Hut (FI).

(22) International Filing Date: 25 February 2004 (25.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
20030294 26 February 2003 (26.02.2003) FI

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(71) Applicant (for all designated States except US):
HELSINKI UNIVERSITY OF TECHNOLOGY
[FI/FI]; P.O. Box 1000, FI-02015 Hut (FI).

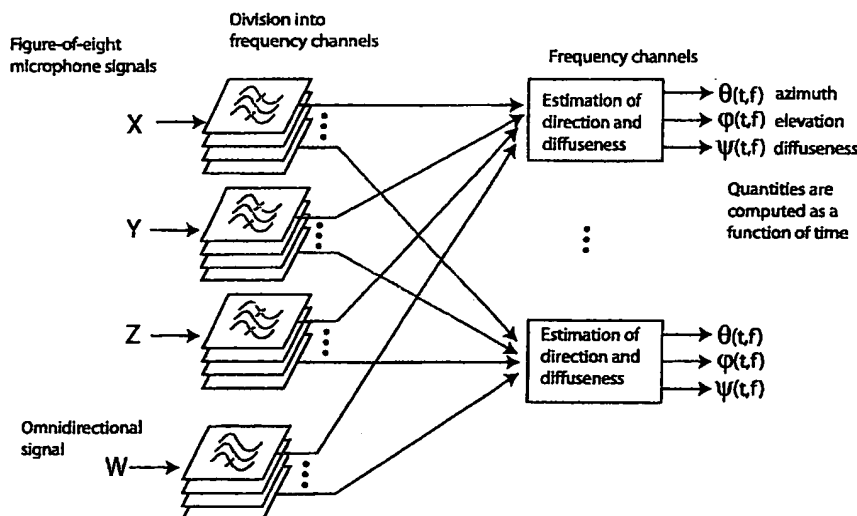
(72) Inventors; and

(75) Inventors/Applicants (for US only): **LOKKI, Tapio**
[FI/FI]; Koskelantie 25 B 17, FI-00610 Helsinki (FI).
MERIMAA, Juha [FI/FI]; Melkonkatu 5 C 44, FI-00210

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), Euro-
pean (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR,
GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK,

[Continued on next page]

(54) Title: A METHOD FOR REPRODUCING NATURAL OR MODIFIED SPATIAL IMPRESSION IN MULTICHANNEL LISTENING



(57) Abstract: The invention concerns a method for reproducing spatial impression of existing spaces in multichannel or binaural listening. It consists of following steps/phases: a) Recording of sound or impulse response of a room using multiple microphones, b) Time- and frequency-dependent processing of impulse responses or recorded sound, c) Processing of sound to multichannel loudspeaker setup in order to reproduce spatial properties of sound as they were in recording room, and (alternative to c), d) Processing of impulse response to multichannel loudspeaker setup, and convolution between rendered responses and an arbitrary monophonic sound signal to introduce the spatial properties of the measurement room to the multichannel reproduction of the arbitrary sound signal, and is applied in sound studio technology, audio broadcasting, and in audio reproduction.



TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

A Method for Reproducing Natural or Modified Spatial Impression in Multichannel Listening

The invention concerns a method for reproducing spatial impression of existing spaces in multichannel or binaural listening. It consists of following steps/phases:

1. Recording of sound or impulse response of a room using multiple microphones,
2. Time- and frequency-dependent processing of impulse responses or recorded sound,
3. Processing of sound to multichannel loudspeaker setup in order to reproduce spatial properties of sound as they were in recording room,
4. (alternative to 3.) Processing of impulse response to multichannel loudspeaker setup, and convolution between rendered responses and an arbitrary monophonic sound signal to introduce the spatial properties of the measurement room to the multichannel reproduction of the arbitrary sound signal,

and is applied in sound studio technology, audio broadcasting, and in audio reproduction.

When listening to sound, a human listener always perceives some kind of a spatial impression. The listener can detect both the direction and the distance of a sound source with certain precision. In a room, the sound of the source evokes a sound field consisting of the sound emanating directly from the source, as well as of reflections and diffraction from the walls and other obstacles in the room. Based on this sound field, the human listener can make approximate deductions about several physical and acoustical properties of the room. One goal of sound technology is to reproduce these spatial attributes as they were in a recording space. Currently, the spatial impression cannot be recorded and reproduced without considerable degradation of quality.

The mechanisms of human hearing are fairly well known. The physiology of the ear determines the frequency resolution of hearing. The wide-band signals arriving at the ears of a listener are analyzed using approximately 40

frequency bands. The perception of spatial impression is mainly based on the interaural time difference (ITD) and interaural level difference (ILD), that are also analyzed within the previously mentioned 40 frequency bands. The ITD and ILD are also called localization cues. In order to reproduce the inherent spatial information of a certain acoustical environment, similar localization cues need to be created during the reproduction of sound.

Consider first loudspeaker systems and the spatial impression that can be created with them. Without special techniques, common two-channel stereophonic setups can only create auditory events on the line connecting the loudspeakers. Sound emanating from other directions cannot be produced. Logically by using more loudspeakers around the listener, more directions can be covered and a more natural spatial impression can be created. The most well known multichannel loudspeaker system and layout is the 5.1 standard (ITU-R 775-1), which consists of five loudspeakers at azimuth angles of 0° , $\pm 30^\circ$ ja $\pm 110^\circ$ with respect to each other. Other systems with varying number of loudspeakers located at different directions have also been proposed. Some existing systems, especially in theaters and sound installations, also include loudspeakers at different heights.

Several different recording methods have been designed for the previously mentioned loudspeaker systems, in order to reproduce the spatial impression in the listening situation as it would be perceived in the recording environment. The ideal way to record spatial sound for a chosen multichannel loudspeaker system would be to use the same number of microphones as there are loudspeakers. In such a case, the directivity patterns of the microphones should also correspond to the loudspeaker layout such that sound from any single direction would only be recorded with one, two, or three microphones. The more loudspeakers are used, the narrower directivity patterns are thus needed. However, current microphone technology cannot produce as directional microphones as would be needed. Furthermore, using several microphones with too broad directivity patterns results in a colored and blurred auditory perception, due to the fact that sound emanating from a single direction is always reproduced with a greater number of loudspeakers than necessary. Hence, current microphones are best suited for two-channel

recording and reproduction without the goal of a surrounding spatial impression.

The problem is, how to record spatial sound to be reproduced with varying multichannel loudspeaker systems.

If the microphones are placed close to sound sources, the acoustics of the recording room have little effect on the recorded signals. In such a case, the spatial impression is added or created with reverberators while mixing the sound. If the sound is supposed to produce a perception as if it were recorded in a specific acoustical environment, the acoustics can be simulated by measuring a multichannel impulse response and convolving it with the source signal using a reverberator. This method produces loudspeaker signals that correspond to recording the sound source in the acoustical environment where the impulse responses were measured. The problem is then, how to create appropriate impulse responses for the reverberator.

The invention is a general method for reproducing the acoustics of any room or acoustical environment using an arbitrary multichannel loudspeaker system. This method produces a sharper and more natural spatial impression than can be achieved with existing methods. The method also enables improvement of the acquired acoustics by modifying certain room acoustical parameters.

Earlier methods

As pertaining to multichannel loudspeaker systems, spatial impression has earlier been created with ad hoc methods invented by professional sound engineers. These methods include utilization of several reverberators and mixing the sound recorded with microphones placed both close to and far away from sound sources in the recording environment. Such methods cannot accurately reproduce any specific acoustical environment, and the final result may sound artificial. Furthermore, the sound always needs to be mixed for a chosen loudspeaker setup and it cannot be directly converted to be reproduced with a different loudspeaker system.

Two main principles for recording spatial sound have been proposed in the literature, see, e.g. [1].

The first principle utilizes one microphone per each loudspeaker in the reproduction system with intermicrophone distances of more than 10 cm. Some related problems have already been discussed. This kind of techniques create good overall spatial impression, but the perceived directions of the reproduced sound events are vague and their sound may be colored. When using a large number of loudspeakers, it is nearly impossible to use as many microphones in the recording situation. Furthermore, the loudspeaker setup has to be known precisely in advance, and the recorded sound cannot be reproduced with different loudspeaker setups or reproduction systems.

The second group of methods applies directional microphones positioned as close to each other as possible. There are two commercial microphone systems, known as the SoundField and Microflown microphones, that are specifically designed for recording spatial sound. These systems can record an omnidirectional response (W) and three directional responses (X, Y, Z) with figure-of-eight directivity patterns aligned in the directions of the corresponding cartesian coordinate axes. Using these responses, it is possible to create "virtual microphone signals" corresponding to any first-order differential directivity pattern (figure-of-eight, cardioid, hypercardioid, etc.) pointing at any direction.

Ambisonics technology is based on using such virtual microphones. Sound is recorded with a SoundField microphone or an equivalent system, and during reproduction, one virtual microphone is directed towards each loudspeaker. The signals of these virtual microphones are fed to the corresponding loudspeakers. Since first-order directivity patterns are broad, sound emanating from any distinct direction is always reproduced with almost all loudspeakers. Thus, there is plenty of cross-talk between the loudspeaker channels. Consequently, the listening area where the best spatial impression can be

perceived is small, and the directions of the perceived auditory events are vague and their sound is colored.

The invention

The purpose of the invention is to reproduce the spatial impression of an existing acoustical environment as precisely as possible using a multichannel loudspeaker system. Within the chosen environment, responses (continuous sound or impulse responses) are measured with an omnidirectional microphone (W) and with a set of microphones that enables to measure the direction-of-arrival of sound. A common method is to apply three figure-of-eight microphones (X,Y,Z) aligned with the corresponding cartesian coordinate axes. The most practical way to do this is to use a SoundField or a Microflown system, which directly yield all the desired responses.

In the proposed method, the only sound signal fed to the loudspeakers is the omnidirectional response W. Additional responses are used as data to steer W to some or all loudspeakers depending on time.

In the invention, the acquired signals are divided into frequency bands, e.g., using a resolution of the human hearing or better. This can be realized, e.g., with a filterbank or by using short-time Fourier transform. Within each frequency band, the direction of arrival of the sound is determined as a function of time. Determination is based on some standard method, such as estimation of sound intensity, or some cross-correlation-based method [2]. Based on this information, the omnidirectional response is positioned to the estimated direction. Positioning here denotes methods to place a monophonic sound to some direction regarding to the listener. Such methods are, e.g., pair- or triplet-wise amplitude panning [3], Ambisonics [4], Wave Field Synthesis [5] and binaural processing [6].

With such processing it can be assumed that at each time instant at each frequency band similar localization cues are conveyed to the listener as would appear in the recording space. Thus, the problem of too wide microphone

beams is overcome. The method effectively narrows the beams according to the reproduction system.

The method, as described previously, is nevertheless not good enough. It assumes that the sound is always emanating from a distinct direction. This is not the case for example in diffuse reverberation. In the invention, this is solved by estimating at each frequency band at each time instant also the diffuseness of sound, in addition to the direction of arrival. If the diffuseness is high, a different spatialization method is used to create a diffuse impression. If the direction of sound is estimated using sound intensity, the diffuseness can be derived from the ratio of the magnitude of the active intensity to the sound power. When the calculated coefficient is close to zero, the diffuseness is high. Correspondingly, when the coefficient is close to one, the sound has a clear direction of arrival. Diffuse spatialization can be realized by conveying the processed sound to more loudspeakers at a time, and possibly by altering the phase of sound in different loudspeakers.

The following describes the invention as a list. In this case, the method to compute sound direction is based on sound intensity measurement, and positioning is performed with pair- or triplet-wise amplitude panning. Steps 1-4 are referring to Figure 1 and steps 5-7 to Figure 2.

1 The impulse response of an acoustical environment is measured or simulated, or continuous sound is recorded in an acoustical environment using one omnidirectional microphone (W) and a microphone system yielding the signals of three figure-of-eight microphones (X,Y,Z) aligned at the directions of the corresponding cartesian coordinate axes. This can be realized, for instance, using a SoundField microphone.

2 The acquired responses or sound are divided into frequency bands, e.g., according to the resolution of human hearing.

3 At each frequency band, the active intensity of sound is estimated as a function of time.

4 The diffuseness of sound at each time instant is estimated based on the ratio of the magnitude of the active intensity and the sound power. Sound power is derived from the signal W .

5 At each time instant, the signal of each frequency band is panned to the direction determined by the active intensity vector.

6 If the diffuseness at a frequency band at a certain time instant is high, the corresponding part of the sound signal W is panned simultaneously to several directions.

7 The frequency bands of each loudspeaker channel at each time instant are combined, resulting in a multichannel impulse response or a multichannel recording.

The result can be listened to using the multichannel loudspeaker system that the panning was performed for. If an impulse response was processed, the resulting responses can be used in a convolution based reverberator to yield a spatial impression corresponding to that perceived in the recording space. Compared to Ambisonics, the invention provides several advantages:

1 Since a distinctly localizable sound event is always reproduced at most with two or three loudspeakers (in pair- and triplet-wise amplitude panning, respectively), the perceived spatial impression is sharper and less dependent on the listening position in a reproduction room.

2 For the same reason, the sound is less colored.

3 Only one high quality omnidirectional microphone is needed to acquire a high quality multichannel impulse response. The requirements for the microphones used in the intensity measurement are not as high.

The same advantages apply compared to the method using the same number of microphones and loudspeakers in sound recording and reproduction. Additionally:

4 From the data resulting from a single measurement it is possible to derive a multichannel response for an arbitrary loudspeaker system.

When processing impulse responses, the method also provides means to alter the produced reverberation. Most existing room acoustical parameters describe the time-frequency properties of measured impulse responses. These parameters can be easily modified by time-frequency dependent weighting during the reconstruction of a multichannel impulse response. Additionally, the amount of sound energy emanating from different directions can be adjusted, and the orientation of the sound field can be changed. Furthermore, the time delay between the direct sound and the first reflection (in reverberation terms pre-delay) can be customized according to the needs of current application.

Other application areas

A method according to the invention can also be applied to audio coding of multichannel sound. Instead of several audio channels, only one channel and some side information are transmitted. Christof Faller and Frank Baumgarte [7, 8] have proposed a less advanced coding method that is based on analyzing the localization cues from a multichannel signal. In audio coding applications, the processing method produces a somewhat reduced quality compared to the reverberation application, unless the directional accuracy is deliberately compromised. Nevertheless, especially in video and teleconferencing applications the method can be used to record and transmit spatial sound.

Operation

It has been shown that in sound reproduction amplitude panning produces better ITD and ILD cues than Ambisonics [9]. Amplitude panning has for a long time been a standard method for positioning a non-reverberant sound source in a chosen point between loudspeakers. A method according to the

invention improves the reproduction accuracy of a whole acoustical environment.

The performance of the proposed system has been evaluated in formal listening tests using a 16-channel loudspeaker system including loudspeakers above the listener, as well as using a 5.1 setup. Compared to Ambisonics, the spatial impression is more precise and the sound is less colored. The spatial impression is close to the measured acoustical environment.

Loudspeaker reproduction of the acoustics of a concert hall using the proposed method has also been compared to binaural headphone reproduction of recordings made with a dummy head in the same hall. Binaural recording is the best known method to reproduce the acoustics of an existing space. However, high quality reproduction of binaural recordings can only be realized with headphones. Based on comments of professional listeners, the spatial impression was in both cases nearly the same, but in the loudspeaker reproduction the sound was better externalized.

The **detailed realization** of the invention is illustrated with the following example:

1 The impulse responses of the Finnish Oopperatalo or any other performance space are measured such that the sound source is located at three positions on the stage and the microphone system at three positions in the audience area = 9 responses. Equipment: standard PC; multichannel sound card, e.g. MOTU 818; measurement software, e.g. Cool Edit pro or WinMLS; microphone system, e.g. SoundField SPSS 422B.

2 The loudspeaker system for reproduction is defined, for instance 5.1 standard without the middle loudspeaker. In this example the middle loudspeaker is left out because the reverberation is reproduced with a four-channel reverberator.

3 With a software accordant with the invention, impulse responses are computed for all loudspeakers corresponding to each source-microphone combination.

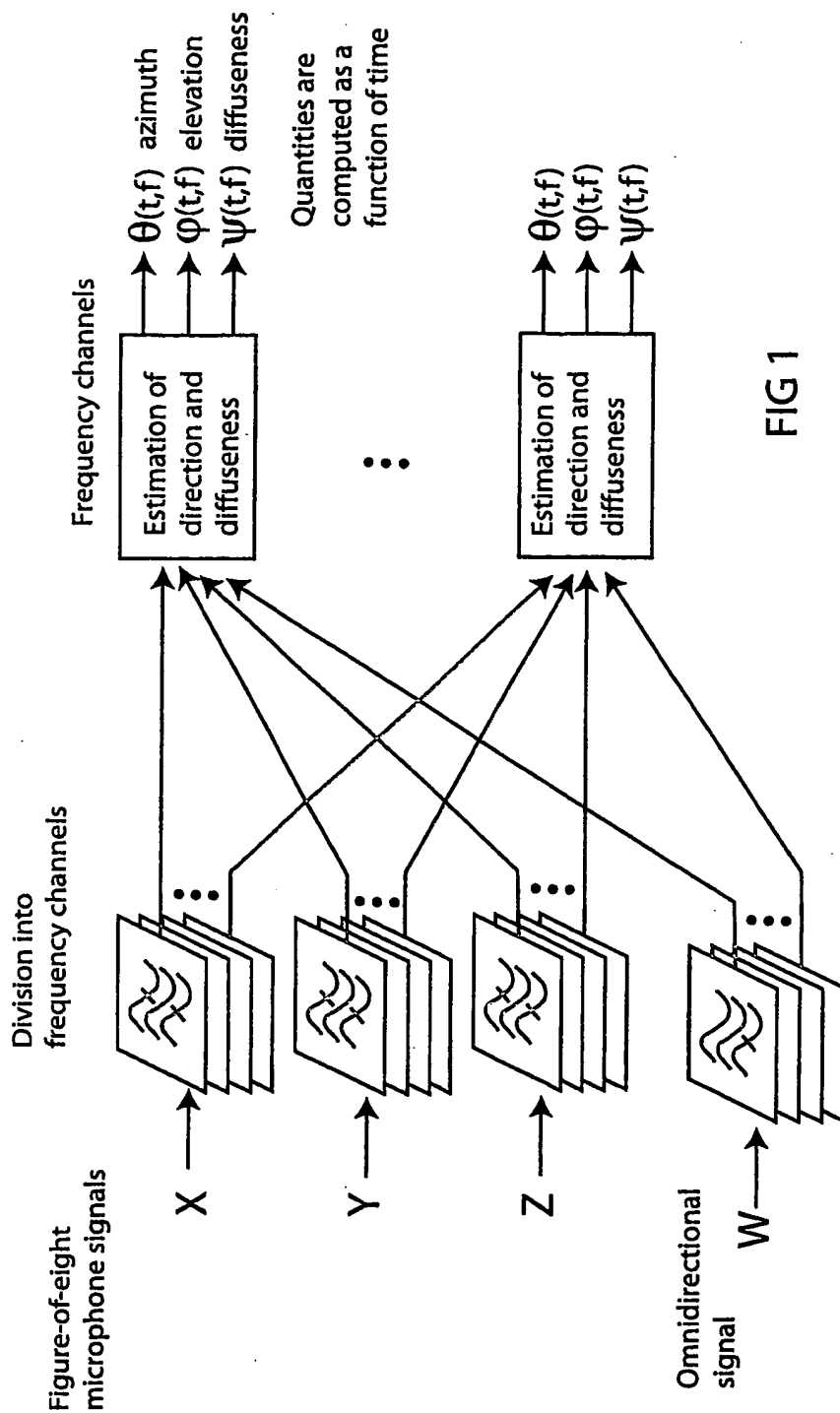
4 Desired source material is convolved with the impulse responses corresponding to one source-microphone combination and the resulting sound is assessed. The sound impression of different source-microphone combinations can be compared in order to choose the one most suitable for current application. Additionally, using several source positions, different source material can be positioned at different locations in the sound field. Equipment can consist of a standard PC or of a convolving reverberator, e.g. Yamaha SREV1; in this case additionally four loudspeakers.

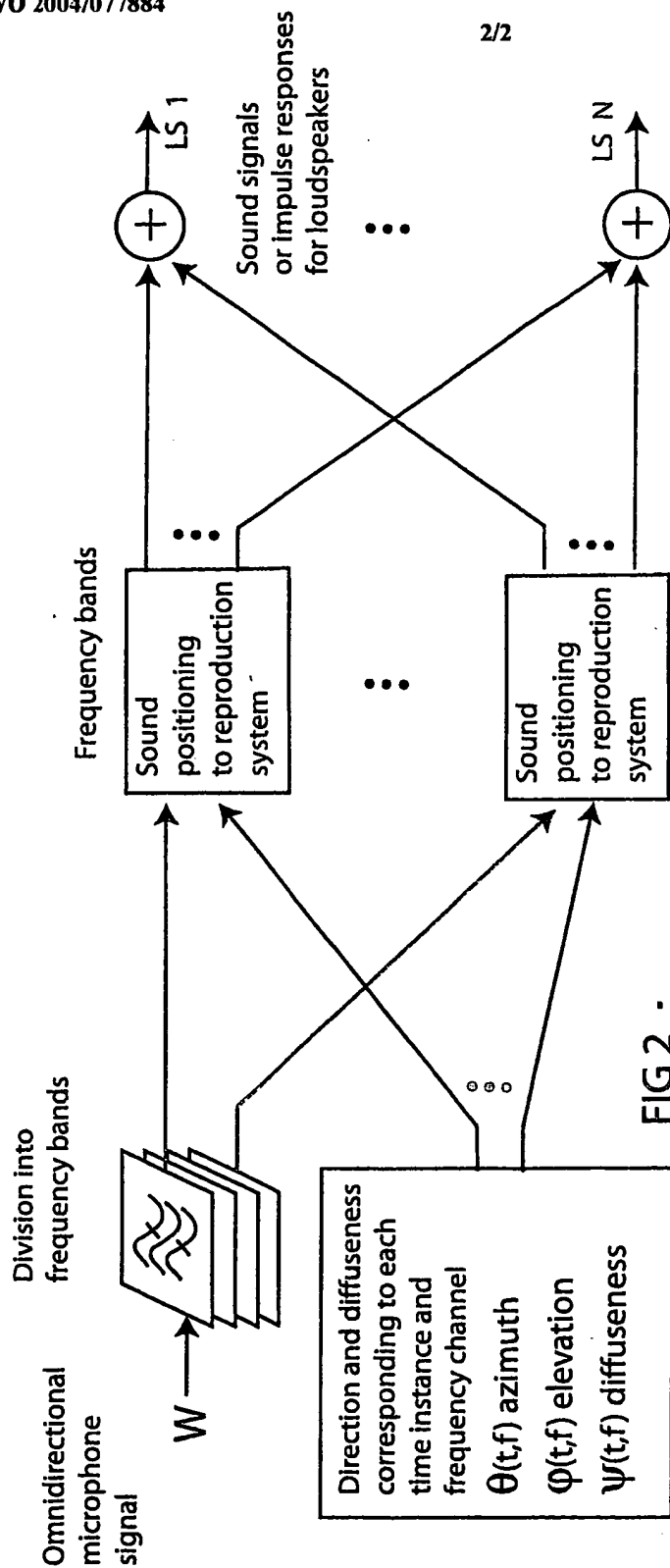
References:

- [1] Farina, A. & Ayalon, R. Recording concert hall acoustics for posterity. AES 24th International Conference on Multichannel Audio.
- [2] Merimaa J. Applications of a 3-D microphone array. AES 112th Conv. Munich, Germany, May 10-13, 2002. Preprint 5501.
- [3] Pulkki V. Localization of amplitude-panned virtual sources II: Two- and three-dimensional panning. J. Audio Eng. Soc. Vol. 49, no 9, pp. 753-767. 2001.
- [4] Gerzon M. A. Periphony: With-height sound reproduction. J. Audio Eng. Soc. Vol 21, no 1, pp. 2-10. 1973
- [5] Berkhout A. J. A wavefield approach to multichannel sound. AES 104th Conv. Amsterdam, The Netherlands, May 16-19, 1998. Preprint 4749.
- [6] Begault D. R. 3-D sound for virtual reality and multimedia. Academic Press, Cambridge, MA. 1994.
- [7] Faller C. & Baumgarte, F. Efficient representation of spatial audio using perceptual parameterization: IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust., New Paltz, USA, Oct. 21-24, 2001.
- [8] Faller C. & Baumgarte, F. Binaural cue coding applied to stereo and multichannel audio compression. AES 112th Conv. Munich, Germany, May 10-13, 2002. Preprint 5574.
- [9] Pulkki, V. Microphone techniques and directional quality of sound reproduction. AES 112th Conv. Munich, Germany, May 10-13, 2002. Preprint 5500.

Claims

1. A method for creating natural or modified spatial impression in multichannel listening, where
 - a) the impulse response of an acoustical environment is measured or continuous sound is recorded using multiple microphones: one omnidirectional microphone (W) and multiple directional or omnidirectional microphones
 - b) the microphone signals are divided into frequency bands according to the frequency resolution of human hearing; and
 - c) based on the microphone signals the direction of arrival and optionally diffuseness of sound is determined at each frequency band at each time instant,**wherein** each frequency channel of an omnidirectional microphone signal is positioned in multichannel listening as a function of time to the direction defined by the estimated direction of arrival of the sound.
2. A method according to claim 1, **wherein** the frequency bands and time instants of the omnidirectional signal W corresponding to non-zero diffuseness are positioned simultaneously to two or more directions in order to create a spatial impression similar to a real acoustical space.
3. A method according to claim 2, **wherein** two or more decorrelated versions of the omnidirectional signal W are created and reproduced simultaneously from two or more directions at frequency bands and time instants corresponding to high diffuseness.
4. A method according to some previous claim, **wherein** the frequency bands applied to each loudspeaker channel are combined in order to produce an impulse response or sound signal for each loudspeaker channel.
5. A method according to some previous claim, **wherein** the processed impulse responses or parts of them are used to produce reverberation with convolution or by modeling them with digital filters.





INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 2004/000093

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: H04S 3/00, H04R 1/20 // H04R 3/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: H04R, H04S

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-INTERNAL, WPI DATA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0869697 A2 (LUCENT TECHNOLOGY INC), 7 October 1998 (07.10.1998) ---	1-5
A	GB 2373956 A (1... LIMITED, INC), 2 October 2002 (02.10.2002) --	1-5
A	US 5812674 A (JOT, J M ET AL), 22 Sept 1998 (22.09.1998) ---	1-5
A	US 683923 B1 (MARASH, J), 19 November 2002 (19.11.2002) -- -----	1-5

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"B" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

2 July 2004

Date of mailing of the international search report

05-07-2004

Name and mailing address of the ISA/

Swedish Patent Office

Box 5055, S-102 42 STOCKHOLM

Facsimile No. +46 8 666 02 86

Authorized officer

Henrik Andersson /LR

Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

Information on patent family members

30/04/2004

International application No.

PCT/FI 2004/000093

EP	0869697	A2	07/10/1998	DE	69801785 D,T	23/05/2002
				JP	10285688 A	23/10/1998
				US	6041127 A	21/03/2000
<hr/>						
GB	2373956	A	02/10/2002	EP	1402755 A	31/03/2004
				GB	0107699 D	00/00/0000
				GB	0207219 D	00/00/0000
				GB	2376595 A,B	18/12/2002
				WO	02078388 A	03/10/2002
<hr/>						
US	5812674	A	22/09/1998	DE	19634155 A	27/02/1997
				FR	2738099 A,B	28/02/1997
				GB	2305092 A,B	26/03/1997
				GB	9617477 D	00/00/0000
<hr/>						
US	683923	B1	19/11/2002	NONE		
<hr/>						

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER: _____**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.